

## A Comparison of Multiple Wearable Technology Devices Heart Rate and Step Count Measurements During Free Motion and Treadmill Based Measurements

Jeffrey Montes<sup>1\*</sup>, Richard Tandy<sup>1</sup>, John Young<sup>1</sup>, Szu-Ping Lee<sup>2</sup>, James Navalta<sup>1</sup>

<sup>1</sup>Department of Kinesiology and Nutrition Sciences, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy. Box 3034, Las Vegas Nevada, 89154, USA

<sup>2</sup>Department of Physical Therapy, University of Nevada, Las Vegas, 4505 S. Maryland Pkwy. Box 3034, Las Vegas Nevada, 89154, USA

**Corresponding Author:** Jeffrey Montes, E-mail: jeffmontes00@gmail.com

### ARTICLE INFO

#### Article history

Received: January 26, 2019

Accepted: March 18, 2019

Published: April 30, 2019

Volume: 7 Issue: 2

Conflicts of interest: None

Funding: None

### ABSTRACT

**Introduction:** Wearable Technology Devices are used to promote physical activity. It is unknown whether different devices measure heart rate and step count consistently during walking or jogging in a free motion setting and on a treadmill. **Purpose:** To compare heart rate and step count values for the Samsung Gear 2, FitBit Surge, Polar A360, Garmin Vivosmart HR+, Scosche Rhythm+ and the Leaf Health Tracker in walking and jogging activities. **Methods:** Forty volunteers participated. Devices were worn simultaneously in randomized configurations. 5-minute intervals of walking and jogging were completed in free motion and treadmill settings with matching paces. Heart rates at minutes 3, 4, and 5 were averaged for the devices along with the criterion measure, the Polar T31 monitor. Step count criterion measure was the mean of two manual counters. A 2x6 (environment vs device) repeated measures ANOVA with Bonferroni post-hoc was performed with significance set at  $p < 0.05$ . **Results:** There was no significant interaction or main effects for walking heart rate. Jogging heart rate saw significant environment and device main effects. Walking step count had a significant interaction between the devices and the environment. Jogging step count had a significant device main effect. **Conclusions:** There may be some conditions such as heart rate measurements taken while walking or step count measurements taken while jogging/running that may only require treadmill-based validity testing.

**Key words:** Heart Rate Determination, Movement, Human Activities, Wearable Electronic Device

### INTRODUCTION

The use of wearable technology devices for obtaining, tracking, and maintaining a healthy life style is becoming more prevalent every year. The number of units sold globally has risen from approximately 23 million in 2014 to 124 million in 2018 (Statista, 2018a). In the same time period, revenue from sales has grown from \$16.7 to \$26.4 billion. It is estimated that by 2022, sales will be in excess of \$73 billion (Statista, 2018b). Because of the influx in types products that can be purchased (watches, bands, bras etc.), consumer interest (Stahl, An, Dinkel, Noble, & Lee, 2016), potential clinical usage (Georgiou et al., 2018; Kisilevsky & Brown, 2016), and the financial investment related to these devices (Coughlin & Stewart, 2016), validated research is required to ensure they are accurate and consistent under a variety of conditions.

One of the issues with wearable technology validation is a lack of standardized testing protocols (Bunn, Navalta, Fontaine, & Reece, 2018). While specific protocols have been proposed by the Consumer Technology Association for validating heart rate (Consumer Technology Association, 2018) and step count measurements (Consumer Technology

Association, 2016), these guidelines have not been officially recognized as the standards by which devices should be tested. Consequentially, researchers have used a variety of methodologies to establish device validity. For heart rate, protocols involving resistance training and cycling (Boudreaux et al., 2018), treadmill walking (Montes, Young, Tandy, & Navalta, 2018), separately evaluated indoor and outdoor free motion walking (Lamont, Daniel, Payne, & Brauer, 2018), and measurements taken while seated, supine, during treadmill walking and running, and when cycling (Wallen, Gomersall, Keating, Wisloff, & Coombes, 2016) have been utilized. For step count, protocols have looked at values compared to a predetermined number of steps (El-Amrawy & Nounou, 2015), steps taken in a predetermined distance (Floegel, Florez-Pregonero, Hekler, & Buman, 2017), values from walking up and down stairs (Huang, Xu, Yu, & Shull, 2016), and treadmill walking (Montes, Young, Tandy, & Navalta, 2017). As presented, a variety of activities and settings have been used. A targeted review of previous research shows free motion walking and jogging and treadmill walking and jogging to be the most commonly used testing protocols.

One of the questions that has been insufficiently addressed is whether there is a difference between values measured during free motion and treadmill-based activities. Most current validity testing utilizes a treadmill under laboratory conditions (Dondzila, Lewis, Lopez, & Parker, 2018). This mode represents a convenient way to administer the test for both researchers and participants, allows for the control of the testing environment, and does not require approval from non-institution-based entities to use off campus facilities (i.e. City and National Parks, Bureau of Land Management etc.). However, the generalization of results from a treadmill or laboratory to a free motion setting may not be practical (Kooiman et al., 2015). In a free motion setting a participant’s speed and intensity can decrease towards the end of a protocol due to fatigue, changes in course direction and elevation can affect values, natural obstacles or other people can interfere, and both the free motion and/or treadmill-laboratory testing may cause anxiety or discomfort for some depending on the setting involved.

The purpose of this research is: 1) to determine if there is a significant interaction between the testing environment and the devices for both heart rate and step count measurements when free motion walking is compared to treadmill walking and when free motion jogging is compared to treadmill jogging. If there is no significant interaction, 2) to determine if there is a significant environment main effect for heart rate and step count measurements when free motion walking is compared to treadmill walking and when free motion jogging is compared to treadmill jogging, and 3) to determine if there is a significant device main effect for heart rate and step count measurements when free motion walking is compared to treadmill walking and when free motion jogging is compared to treadmill jogging. To date, we are unaware of any research that has specifically looked at these comparisons. We hypothesized that: 1) there would be no significant interaction between the environment and the devices for heart rate and step count measurements when free motion and treadmill activities were compared to one another, 2) there would be no significant environment main effect, and 3) there would be no significant device main effect.

**METHODS**

**Participants and Design of Study**

This study utilized a cross-sectional, repeated measures research design investigating the differences in recorded heart rate and step count values for five wearable technology devices in different applications. Free motion and treadmill walking were compared to one another as was free motion and treadmill jogging. Participants attended two data collection sessions. The first was to record free motion walking and jogging values during 5-minute walking intervals. The second

was to record the same but on a treadmill. The purpose was to evaluate if there was a difference in recorded values between the two conditions for each motion. Forty healthy (identified as low risk according to the ACSM pre-participation screening questionnaire) participants aged 25.09±7.17 years (twenty males and twenty females) volunteered for this investigation (descriptive characteristics are provided in Table 1.). Participants filled out an informed consent form that was approved by the UNLV Biomedical Institutional Review Board (#885569-3). At the time of this study, there was no previous research data to calculate a definitive “N” size. The use of forty participants was agreed upon by all contributing authors. This was based on previous but not standardized recommendations by the Consumer Technology Association to use twenty participants for both heart rate and step count testing purposes (Consumer Technology Association, 2016, 2018). To be conservative, this value was doubled.

**Devices**

The six wearable technology devices investigated consisted of four that are worn on the wrist: the Samsung Gear 2, FitBit Surge, Polar A360, and the Garmin Vivosmart HR+, one worn on the waist: Leaf Health Tracker, and one is worn on the upper forearm: Scosche Rhythm+. Five of the devices measured heart rate: the Samsung Gear 2, FitBit Surge, Polar A360, Garmin Vivosmart HR+, and the Scosche Rhythm+. The chest mounted Polar T31 (Lake Success, NY) was used as the criterion measure for heart rate. Five of the devices measured step count: the Samsung Gear 2, FitBit Surge, Polar A360, Garmin Vivosmart HR+, and the Leaf Health Tracker. The average of two manual step counts using a hand-held tally counter (Horsky, New York, NY) was used as the criterion measurement for this measurement. Immediately prior to testing, the participants age, sex, height, weight, and where the device was being worn were programmed into each device. The device was synchronized, and the appropriate “activity” mode, if available, was selected. All devices that measured heart rate used proprietary green wavelength LED photoplethysmography. All devices that recorded step count used proprietary algorithms to determine what constitutes a step for counting purposes.

The Samsung Gear 2 (Samsung Electro-Mechanics, Seoul, South Korea) is a wrist-worn smartwatch. Sensors include an accelerometer, gyroscope, and heart rate monitor.

The Fitbit Surge (Fitbit Inc, San Francisco, CA) is a fitness super wrist-watch that utilizes GPS tracking to determine distance and pace. Sensors and components include 3-axis accelerometers, digital compass, optical heart rate monitor, altimeter, ambient light sensor, and vibration motor.

The Polar A360 (Polar Electro, Kempele, Finland) is a wrist-worn fitness tracker that has a proprietary optical heart rate module. No other specifications are given.

**Table 1.** Participants characteristics. Means ± SD presented

	<b>Age (yrs)</b>	<b>Height (cm)</b>	<b>Mass (kg)</b>	<b>BMI (m/kg<sup>2</sup>)</b>
All participants (N=40)	25.09±7.17	169.64±11.18	77.19±19.2	26.43±5.19

BMI = Body Mass Index

The Garmin Vivosmart HR+ (Garmin Ltd, Canton of Schaffhausen, Switzerland) is smart activity tracker with wrist-based heart rate as well as GPS. Sensors include a barometric altimeter and accelerometer.

The Rhythm+ (Scosche Industries, Oxnard, CA) is a forearm-based heart rate tracker that uses an optional green or yellow LED colored PPG sensor. Unlike the wrist-worn devices, it does not have a display window. It uses a third-party application downloaded to a smartphone or tablet to show HR measurements. This study used the MotiFIT application (version 1.3.4(56), Dieppe, New Brunswick, CANADA) on a Samsung Galaxy S8+ smartphone (Samsung, Ridgefield Park, NJ).

Leaf Health Tracker (LF; Bellabeat, San Francisco, CA): Waist worn sensor. Include a 3-axis accelerometer and vibration motor.

### Protocol

Data for this study was completed concurrently during a collection period that has been recently published (Montes & Navalta, 2019). The protocol has been repeated here for the convenience of the reader. In the week prior to testing, participants provided anthropometric data. Age in years was self-reported, height (cm) was measured with a Health-o-meter wall mounted height rod (Pelstar LLC/Health-o-meter, McCook, IL), mass (kg) and Body Mass Index (BMI) was provided by a hand-and-foot bioelectric impedance analyzer (seca mBCA 514 Medical Body Composition Analyzer, Seca North America, Chino, CA).

On the first day of testing, participants were fitted with the Samsung Gear 2, FitBit Surge, Polar A360, Garmin Vivosmart HR+, Scosche Rhythm+ and Leaf Health Tracker. They then proceeded to a long indoor hallway with cones spaced 200 feet apart. Participants sat for 5 minutes and then completed the first 5-minute self-paced free motion walk back and forth between the cones. Participant heart rate was recorded for minutes 3, 4, and 5 while step count was recorded by the two manual counters. After a 5-minute seated rest period, participants completed the first 5-minute self-paced free motion jog. Heart rate for minutes 3, 4, and 5 and the step count by two manual counters were again recorded. Participants then rested in a seated position for 10 minutes. They then performed a second self-paced 5-minute free motion walk and jog in the same manner as the first with heart rate and step count recorded in the same manner. The two manual counters for all free-motion walks and jogs were positioned near the center of the testing area but were separated so they could not view each other's thumb motion nor hear the "clicking" from the tally counter. This prevented any synchronized counting between the two. The manual counters were instructed not to follow or move with the participants to prevent influencing their walking/jogging speed. The distance traveled for both free motion walks and jogs was measured and the speed in miles per hour was calculated and rounded to the nearest 0.1.

One to two days later at approximately the same time of day ( $\pm 1$  hour), the participants returned for treadmill-based walking and jogging. They were fitted with all the devices

in the same manner and configuration as on day two. All treadmill activities were performed on a Trackmaster treadmill (Full Vision, Inc. Newton, KS). After a 5-minute seated rest period, they completed the first 5-minute treadmill walk at the speed calculated from the first free motion walk. Participant heart rate was recorded for minutes 3, 4, and 5 with the step count recorded by the two manual counters. Following a 5-minute seated rest period, they completed the first 5-minute treadmill jog at the speed calculated from the first free motion jog. Heart rate for minutes 3, 4, and 5 and the step count by two manual counters was again recorded. Participants rested in a seated position for 10 minutes. They then performed a second 5-minute treadmill walk and jog with the heart rate and step count recorded in the same manner as the first treadmill activities. Speeds for the second treadmill walk and jog were calculated from the second free motion walk and jog. Speeds were replicated on the treadmill in order to normalize the distance a participant traveled in the 5-minute testing intervals for both conditions. The grade for all treadmill testing was set to 0%. The two manual counters were positioned at opposite sides of the lab in order to prevent any synchronized "clicking".

### Statistical Analysis

IBM SPSS (IBM Statistics version 24.0, Armonk, NY) was used for all statistical analysis. Heart rate values for minutes 3, 4, and 5 were averaged together to give one value that represented a steady state heart rate for each device. The tested device values were compared to the Polar T31. Recorded step counts for the tested devices for each 5-minute activity were compared to the mean of two manual step counters. Three outliers of  $\geq \pm 3$  standard deviations were removed from the step count analysis (participant #7 and #14, FitBit Surge, free motion jog: step count was not recorded properly at the end of both said activities. Participant #37, Samsung Gear 2, treadmill walk: device stopped counting and had to be re-synchronized to reset step counting function for next activity). A 2x6 repeated measures ANOVA with Bonferroni post-hoc analyses was performed using two conditions, 1) the free motion and treadmill environment and 2) the six device measurements that included the five tested wearable technology devices and the indicated criterion measure. Mauchly's Test of Sphericity was performed with the Huynh-Feldt adjustment used as the correction factor when required. Significance was set at  $< 0.05$ .

### RESULTS

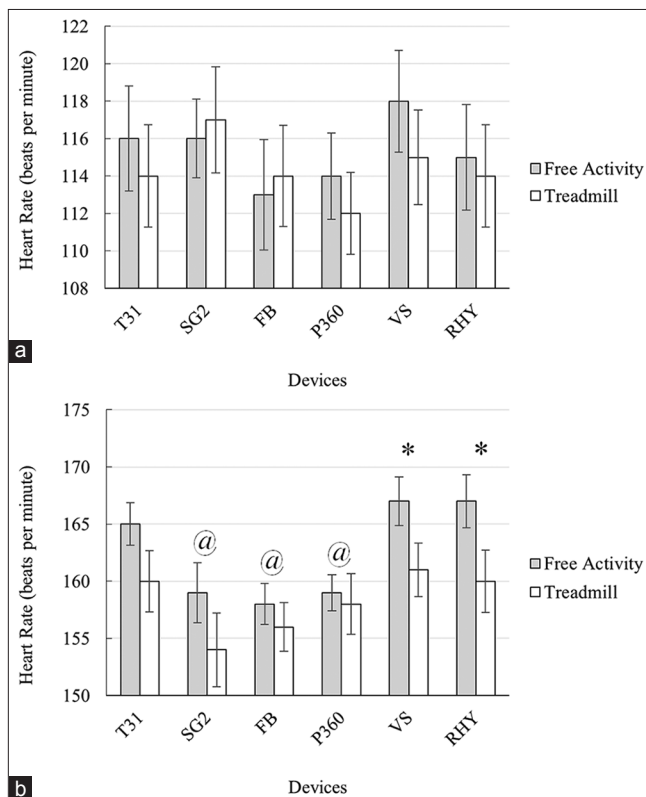
For heart rate measurements compared between free motion and treadmill walking, there was no significant interaction between the environment and the wearable technology devices,  $F(2.81, 109.49)=0.95$ ,  $p=0.416$ , no significant environment main effect,  $F(1, 39)=0.46$ ,  $p=0.502$ , and no significant device main effect,  $F(2.36, 91.86)=1.64$ ,  $p=0.195$  (Figure 1a).

For heart rate measurements that were compared between free motion and treadmill jogging, there was no significant interaction between the environment and the wearable

technology devices,  $F(3.58, 139.79)=2.04$ ,  $p=0.099$ . Both the environment and device main effects were significant,  $F(1, 39)=6.91$ ,  $p=0.012$  and  $F(3.85, 150.27)=9.53$ ,  $p<0.001$  respectively. The Samsung Gear 2 ( $p=0.007$ ), FitBit Surge ( $p=0.016$ ), and the Polar A360 ( $p=0.017$ ) all had significantly lower mean heart rates compared to the Polar T31 (Figure 1b).

For step count measurements compared between free motion and treadmill walking, there was a significant interaction between the environment and the wearable technology devices:  $F(3.86, 146.57)=2.65$ ,  $p=0.037$ . Simple effect analysis indicated that the interaction was due to the effect of one device in the laboratory environment. The Polar A360 returned a significantly greater step count during free motion walking over treadmill walking ( $p=0.020$ ). Simple effect analysis also provided evidence that the Samsung Gear 2 ( $p<0.001$ ), FitBit Surge ( $p<0.001$ ), and the Polar A360 ( $p<0.001$ ) returned significantly lower step counts compared to the manual counters (Figure 2a).

For step count measurements compared between free motion and treadmill jogging, there was no significant interaction between the environment and the wearable technology

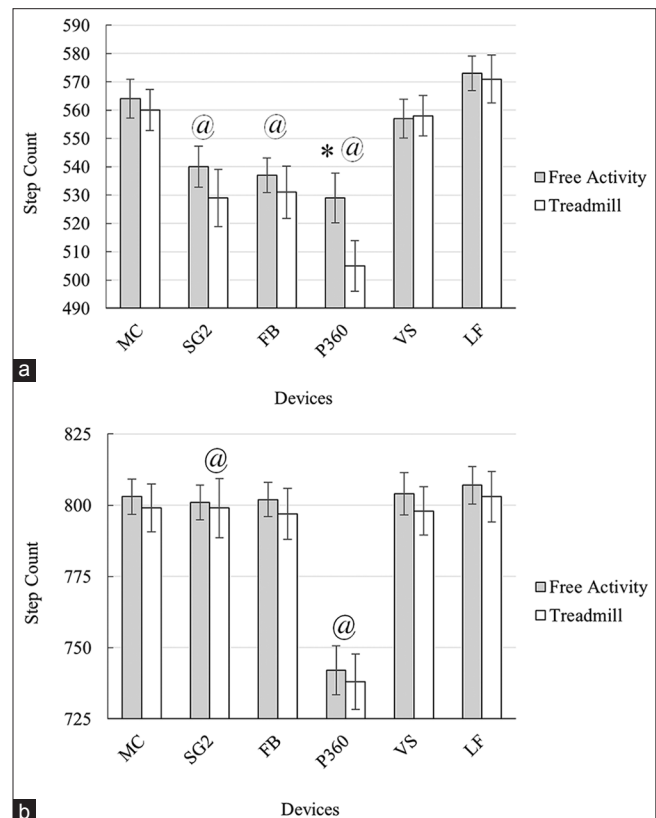


**Figure 1.** (a) Comparison of steady state heart rate average between free motion and treadmill walking. Standard error indicated by error bars. Polar T31=T31, Samsung Gear 2=SG2, FitBit Surge=FB, Polar A360=P360, Garmin Vivosmart HR+=VS, Scosche Rhythm+= RHY. (b) Comparison of steady state heart rate average between free motion and treadmill jogging. Standard error indicated by error bars. \* Indicates a significant difference ( $p<0.05$ ) between the device's free motion and treadmill values. @ Indicates a significant mean difference ( $p<0.05$ ) between the device and the criterion measure. Polar T31=T31, Samsung Gear 2=SG2, FitBit Surge=FB, Polar A360=P360, Garmin Vivosmart HR+=VS, Scosche Rhythm+= RHY

devices,  $F(3.14, 116.18)=2.10$ ,  $p=0.054$  and no significant environment main effect,  $F(1, 37)=1.92$ ,  $p=0.174$ . There was a significant device main effect  $F(1.90, 70.15)=63.12$ ,  $p<0.001$ . The Samsung Gear 2 ( $p=0.007$ ) and the Polar A360 ( $p<0.001$ ) both had significantly lower step counts measurements than the manual counters. (Figure 2b).

## DISCUSSION

The aim of the current study was to evaluate any potential differences between free motion and treadmill environments during walking and jogging for heart rate and step count measurements. We hypothesized that: 1) there would be no significant interaction between the environment and the devices for heart rate and step count measurements when free motion and treadmill activities were compared to one another, 2) there would be no significant environment main effect, and 3) there would be no significant device main effect. To our knowledge, no previous research on wearable technology



**Figure 2.** (a) Comparison of step counts between free motion and treadmill walking. Standard error indicated by error bars. \*Indicates a significant difference ( $p<0.05$ ) between the device free motion and treadmill values. @ Indicates a significant mean difference ( $p<0.05$ ) between the device and the criterion measure. Manual Count=MC, Samsung Gear 2=SG2, FitBit Surge=FB, Polar A360=P360, Garmin Vivosmart HR+=VS, Leaf Health Tracker=LF. (b) Comparison of step counts between free motion and treadmill jogging. Standard error indicated by error bars. @ Indicates a significant mean difference ( $p<0.05$ ) between the device and the criterion measure. Manual Count=MC, Samsung Gear 2=SG2, FitBit Surge=FB, Polar A360=P360, Garmin Vivosmart HR+=VS, Leaf Health Tracker=LF

gy devices has evaluated these comparisons simultaneously. We observed no significant interaction or device or environment main effects for walking heart rate measurements. While measurements for the jogging heart rate did not have a significant interaction, there were significant device and environment main effects. Walking step count produced a significant interaction between the devices and the environment. Jogging step count had only a significant device main effect.

### Heart Rate

Heart rate while walking produced no significant interactions or main effects. For the comparison between free motion and treadmill walking, all the tested devices along with the Polar T31 measured heart rate with statistically similar values. While heart rate measurements during jogging had no significant interaction between the devices and the environment, there were significant main effects due to the environment and significant main effects between the device heart rate values and the Polar T31 criterion measure.

Heart rate values are instantaneous measurements. The primary influence on their value is the intensity of the activity being performed. We extrapolated the treadmill walking and jogging speeds from the corresponding free motion walking and jogging activities. In theory, the effort exerted along with the corresponding heart rates should have been similar for both movements in both settings. This was the case for the walking activities. However, for the jogging activities there were noticeable differences. While treadmill speeds remain constant, free motion speeds can vary depending on the length of the protocol and the fitness level of the participant. Both factors could create a scenario in which the tested individual begins a free motion jogging protocol in a rapid manner but later decrease in speed due to fatigue as they adjust their speed according to the exertion level. When jogging fatigued on a treadmill, participants would be expected to expend more effort to maintain the constant rate of speed required later in a protocol due to the inability to slow down. This inability to slow down on a treadmill, especially at higher speeds, should hypothetically force an increase in exertion, and thus higher heart rates. However, our research offered evidence of the opposite. Overall, the wearable technology devices registered higher heart rate measurements when jogging in a free motion setting than when on a treadmill. The Polar T31, Garmin Vivosmart HR+ and Scosche Rhythm+, all had significantly higher values during free motion jogging. The Samsung Gear 2, Fitbit Surge, and the Polar A360 showed a trend toward increased heart rate in the free motion setting, but the measures were not significant. (Figure 1b). Thus, it would be logical to conclude that there are indeed factors related to the setting that influence this outcome regarding heart rate differences.

With regard to the devices themselves, the Polar T31, unlike the 5 other tested devices, uses its location on the sternum to detect electrical impulses during cardiac contractions to measure heart rate (“How does a Polar Training Computer measure heart rate?,” 2018). The tested wearable technology devices all employ photoplethysmography (PPG). PPG uses LED light that is projected into the underlying

skin surface. The transmitted and reflected light is used to measure the expansion and contraction of near surface blood vessels as they are impacted by pressure waves from a contracting heart (Maeda, Seaman, & Tamura, 2010). However, the wavelength emitted by an LED light can vary greatly (Maeda, Sekine, & Tamura, 2011). Each device utilizes its own proprietary measuring technique that comprises of not only proprietary LED wavelengths but also proprietary algorithms. As a result, it may be that the Scosche Rhythm+ and the Garmin Vivosmart HR+ are determining heart rate measurements with either an appropriate wavelength and/or more precise algorithm.

Another factor to explain increased heart rate during free motion compared to treadmill activity may be that the moving treadmill belt helps with motion, making the activity easier. Walking in general involves overcoming both gravity (vertical motion) and producing enough horizontal force to propel one’s body forward (horizontal motion). While the effect of gravity is relatively similar in either environment, a moving treadmill belt minimizes the force required to move horizontally which keeps exertion levels lower. Our study used a grade of 0% for all treadmill motion. Research has shown that a treadmill grade of approximately 1% induces an exertion equivalent to that of free motion (Jones & Doust, 1996). The self-selected jogging speed, 0% treadmill grade, and the moving belt appear to have been the stronger stimuli resulting in the lower treadmill heart rate measurements. When walking, heart rate values do not seem to be affected as this represents a relatively low intensity exercise. Jogging, however, can be classified as moderate to high intensity depending on one’s fitness level which may have led to more variation in the participant’s heart rate range (Figure 1b) (Liguori, Dwyer, & Fitts, 2014). While our protocol was only for 5-minute intervals, this amount of time appears to have been enough for those in the study to show the effects due to the difference of the two motions.

A psychological aspect that may have influenced the higher free motion jogging heart rate values may have been a result resembling the “white coat” effect that persons normally experience in a medical setting. The white coat effect is loosely defined as differences in heart rate and blood pressure values when measured in a clinical setting or by medical personal versus when taken in a normal or relaxed environment (Pickering, Gerin, & Schwartz, 2002). The assumption being that the presence of a medical professional or a being in clinical setting creates anxiety in the participant, producing higher heart rate and blood pressure measurements than normal (Pickering et al., 2002). Briefly, the setting and the nervousness level of those measured may cause higher readings. In our study, all participants began their testing in a free motion setting that was performed in public. The combination of a public setting and being unfamiliar with the protocol while being observed by the researchers may have contributed to the higher free motion jogging heart rates. Because the treadmill activities were performed one to two days later in a laboratory, the participants were familiarized with the protocol and out of view of the public. Both factors may have reduced any nervousness related to the protocol and lowered heart rate as a result. It must be noted that this

did not seem to affect heart rate while walking as the devices were split between free motion and treadmill recordings for the higher heart rate values.

Previous research on the tested wearable technology devices for heart rate measurements was not consistent with our results. Our results indicated that the Samsung Gear 2 significantly underestimated heart rate when jogging. One separate study showed it had very little difference in heart rate when compared to their unnamed criterion measure when walking (El-Amrawy & Nounou, 2015). Another study indicated that the mean absolute percent error was not acceptable for a variety of activities. This study did not specify if the estimation was higher or lower though (Shcherbina et al., 2017). In our study, the FitBit Surge had significant lower jogging heart rate measurements when compared to the Polar T31. Research on the FitBit Surge by (Thiebaud et al., 2018) indicated a small overestimation for walking treadmill activities up to 3mph and a slight underestimation for jogging speeds greater than that. Additionally, they reported the mean absolute percent error was unacceptable for walking but within agreeable tolerances for jogging. Two additional studies for the FitBit Surge (Shcherbina et al., 2017; Xie et al., 2018) both produced unacceptable mean absolute percent errors for heart rate during several different activities. The Polar A360 in our study significantly underestimated heart rate. There is only one known study for this device. It's results indicate that as exercise intensity increases, both the underestimation of heart rate as well as the mean absolute percent error increase accordingly (Boudreaux et al., 2018). Both the Garmin Vivosmart HR+ and the Scosche Rhythm+ had no significant difference in jogging heart rate when compared to the Polar T31. One study for the Garmin Vivosmart HR+ contradicted ours in that those results indicated that as exercise intensity increase, underestimation of heart rate as well as the mean absolute percent error increases (Boudreaux et al., 2018). Two separate studies on the Scosche Rhythm+ by (Gillinov et al., 2017) and (Stahl et al., 2016) had similar results. Both reported that the Scosche Rhythm+ had minimal bias in measurements and a low mean absolute percent error.

### **Step Count**

A significant interaction between the wearable technology devices and the environment was seen for walking step count measurements. In contrast, jogging step count measurements only presented a significant main effect between the mean values of the devices compared to the manual step count. For all but one condition (free motion walking, Garmin Vivosmart HR+) steps taken while moving in a free motion setting were higher than on a treadmill (Figure 2a, Figure 2b). Wearable technology devices attempt to register each step based on the movement of the body on which the device is placed. Any potential differences in movement patterns between free motion and treadmill activities may result in different results for the same motion. However, previously published literature is not definitive as to what, if any, of these observed differences in motion mechanics may be (Riley et al. 2008; Schache et al., 2001).

Prior research has shown slight differences in certain comprehensive parameters such as stride length and cadence between the two conditions. For example, one study by Murray, Spurr, Sepic, Gardner, & Mollinger (1985) provided evidence that treadmill walking resulted in shorter strides and a quicker cadence while Frishberg (1983) observed no difference when free motion and treadmill walking patterns were compared. Similar to this is the mechanical response of persons to the differences in surfaces they are interacting with. Free motion activities are usually performed on hard surfaces such as asphalt, concrete, or hard rubber. Most treadmills, however, are designed to have a spring effect that returns energy back to the individual (Schache et al., 2001). Studies have shown that walking/jogging over different surfaces results in varying degrees of leg stiffness (Ferris, Louie, & Farley, 1998). These subtle lower extremity adjustments may be supporting the different step count values between the two conditions.

Of the five wearable technology devices tested, the only one that was not wrist worn was the Leaf Health Tracker. For both the walking and jogging step count comparisons, its values were consistently similar to the manual step count. Previous research has shown that device placement does have an influence on step count accuracy. In order to accurately count steps, wearable technology devices need to have high efficiency for the specific areas of the body they are designed for and are placed. A study done by Tudor-Locke, Barreira, & Schuna (2015) compared accuracy levels for wrist worn and waist worn devices with waist worn step counters being more accurate. A limitation to their study, as was in ours, was that different devices were being tested in different body positions. This makes it difficult to confidently compare results to one another. Simpson et al. (2015) compared wearable technology devices worn on the ankle to those worn on the waist. While the ankle position provided slightly more accurate results than that of the waist, both were shown to provide accurate step count values than those recorded by wrist worn devices. A take away from our study and those conducted previously is that device placement on other than the wrist may be preferable for those wishing to accurately monitor daily step counts.

Previous research on the tested wearable technology devices for step count measurements were not consistent with our results. The Samsung Gear 2 significantly underestimated steps in both walking and jogging when compared to a manual count of steps. Only one known study corroborated that result (Modave et al., 2017) while another indicated that it overestimated (El-Amrawy & Nounou, 2015). For the FitBit Surge, our results showed a significant underestimation of steps counted for walking when compared to the manual count but a very small underestimation when jogging. Discrepancies in treadmill walking and jogging for the FitBit Surge were also observed by Binsch, Wabeke, & Valk (2016) and a significant underestimation of steps in free motion walking was recorded by Modave et al. (2017). The Polar A360, significantly underestimated steps when both walking and jogging. In addition, there was a significant main effect from the environment during walking. While there is one known study that corroborates the underestimation of

the step count measurement (Bunn et al., 2018), there is one that reports it overestimates it (Navalta et al., 2018). Both the Garmin Vivosmart HR+ and the Leaf Health Tracker had no significant mean differences between the measured values and the manual count for walking or jogging. For the Garmin Vivosmart HR+, two studies had similar results (Lamont et al., 2018; Wahl, Duking, Droszez, Wahl, & Mester, 2017). The only known study on the Leaf Health Tracker also concurred (Navalta et al., 2018).

### Testing

As discussed previously, there is no consistency in the literature for testing wearable technology devices. This means there is no practical manner to compare the results of one study to another. Resources and time are potentially wasted testing the same wearable technology devices by several researchers with different applications. Consequently, this leads to many varied statistical conclusions due to the different numbers of participants, how and when values are recorded, and the variety of activities that can be utilized. Moreover, in many studies only one distinct value was recorded and analyzed at a time. The use of a commonly accepted protocol that allows for numerous measurements to be taken simultaneously would be the most efficient use of resources and time. Established protocols would also allow for the timely testing of devices as they become available. This is a vital component for wearable technology testing as a plethora of new devices are quickly and continuously being procured by many entities. Consequently, currently available devices are rapidly being replaced or being regulated to obscurity by newer or alternate versions. Many times, they become obsolete before a proper evaluation and reporting of results to the public can be made (Bunn et al., 2018).

To this end, the Consumer Technology Association has procured recommendations regarding standardized testing protocols for both heart rate and step count validation. While these suggested protocols can be viewed as forward thinking, the practicality of the testing methods are not entirely feasible. Their recommendation for heart rate is that it should be recorded at least once every 5 seconds (Consumer Technology Association, 2018). To fulfil this testing standard, specific software and/or hardware that captures heart rate signals from numerous devices simultaneously and subsequently inputs them into a common spread sheet is required. The software and equipment cost may represent aspects that some investigators may not be able to handle due to financial restraints or a lack of suitable technology that supports the said program. They also advocate that step count activities be video recorded with two manual counters separately reviewing the footage at a later time/date. Both counters would have to come up with the exact same count for it to be considered a valid value (Consumer Technology Association, 2016). This is not practical in a free motion setting as camera use may be hindered by visual obscurements, possible changes in elevation and movement direction, and the interference of persons as the participants move through the public testing area. The testing protocol we have utilized for this study employs the average of several heart rates during an activity to

represent a steady state measurement. The idea being that it represents a single value for analysis purposes. Also, our use of two manual step counters allows for flexibility and mobility in almost every environment. This step count method has already been used in previous research with inter-rater reliability being  $\geq 0.99$  for all analysis (Floegel et al., 2017; Navalta et al., 2018).

Our research protocol for this study was unique in that: 1) All persons performed two 5-minute free motion walks and two 5-minute free motion jogs on the same day. 2) One to two days later all persons performed two 5-minute treadmill walks and two 5-minute treadmill jogs at approximately  $\pm 1$  hour as the free motion activities. 3) Because we used the same persons for both days of testing, we were able to reasonably compare the heart rate and step count results of the two settings used for walking and jogging. We feel that this protocol is a sensible and practical way to test wearable technology devices. As it is not confined to just heart rate and step count measurements, energy expenditure, ventilation rate, step cadence, and distance traveled can all be evaluated concurrently as well. This procedure would also allow for simultaneous test-retest and validity analysis.

Low intensity physical activity has been shown to increase the accuracy for devices that use PPG (Maeda et al., 2011). Conversely, high intensity activities such as jogging or running increase the accuracy of devices that record step count (Schneider, Crouter, & Bassett, 2004). Both studies correspond to our results regarding our results and the respective criterion measurements. This means heart rate during jogging and step count during walking may be inaccurate due to factors such as a device's measurement mechanism or because the associated movement from the activity being performed is not within the parameters for accurate recording. While the concept of only using a treadmill was extrapolated from the six devices tested in this study, the potential for the development of future testing standards is exciting. The implication is that minimal validation testing requirements could save time, effort, and resources in future investigations. However, the fact that the jogging heart rate and walking step count measurements had potential influences from the testing environment shows that not all activities may fit the criteria for treadmill specific testing. Because of this conflict in results, there may be no choice but to test future devices not only in the settings we normally utilized but in other less common ones such as hiking or in mimicking daily life activities. However, if device testing using only a treadmill in a controlled setting can be proven to be adequate, the benefits from this development would be highly advantageous. There would be minimal interference while observing participants, heart rate monitors could be supervised with ease, and if video recording is required, it would be easy to do so.

One factor that was not controlled for nor was recognized until after the data collection was complete was the potential effect of the ambient temperature during both conditions. The free motion activities were conducted in an interior building hallway while the treadmill activities were performed in a controlled laboratory setting. Temperatures were not recorded for either. However, the laboratory setting

utilized for this study is normally cooler than the building hallway areas. Body temperatures may have been higher in the free motion setting due to the higher temperatures in that environment. This may have resulted in greater dilation of blood vessels for the dissipation of body heat. This would result in the heart pumping faster to maintain blood pressure (Wilson & Crandell, 2011).

A limitation of this study was that most of the participants were young, healthy college students (age 25.09±7.17) with slightly above normal body mass index values (26.43±5.19). It has been shown that factors such as high or low body fat composition (Crouter, Schneider, & Bassett Jr., 2005) or body mass index (Shepherd, Toloza, McClung, & Schmalzried, 1999) may affect the ability of a wearable technology device to accurately record values, especially for steps taken. Additionally, special populations such as the elderly or those who are obese have been shown to have different gait mechanics and physiological factors that may add to the complications regarding the recording of a correct value (Melanson et al., 2004). As such, the application of our results to these other population should be done carefully (Bassett, Rowlands, & Trost, 2012). Future research should concentrate on these various participant populations in order to further determine what specific factors may play a role in a device's accuracy. Correction factors may have to be extrapolated and applied as needed (Wahl et al. 2017).

## CONCLUSION

Our research produced a mixture of results that were dependent on the combination of the type of the device worn and the setting it was utilized in. The recorded difference in values between free motion and treadmill walking and jogging will require further evaluation of lower body gait mechanics, arm swing and its related motion artifact, and the fitness levels of the participants in order to fully understand what factors may or may not significantly contribute to the differences procured. In terms of actual device testing, we addressed to two current issues in this area. First, we introduced a protocol that accounts for the lack of reliability testing that has been observed in much of the literature. Secondly, our protocol shows that testing can be done with a minimal use of required technology. Outside testing may have the element of being more difficult to perform due to physical obstacles, location, weather, and equipment complications. But if it can be shown that treadmill testing can be used in lieu whenever possible, the savings in time and resources would be beneficial to all.

## REFERENCES

- Bassett, D. R., Jr., Rowlands, A., & Trost, S. G. (2012). Calibration and validation of wearable monitors. *Medicine & Science in Sports & Exercise*, 44(1 Suppl 1), S32-38.
- Binsch, O., Wabeke, T., & Valk, P. J. (2016). *Comparison of three different physiological wristband sensor systems and their applicability for resilience- and work load monitoring*. Paper presented at the IEEE 13<sup>th</sup> International Conference on Wearable and Implantable Body Sensor Networks, San Francisco CA.
- Boudreaux, B. D., Hebert, E. P., Hollander, D. B., Williams, B. M., Cormier, C. L., Naquin, M. R., Kraemer, R. R. (2018). Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. *Medicine and Science in Sports and Exercise*, 50(3), 624-633.
- Bunn, J. A., Navalta, J. W., Fountaine, C. J., & Reece, J. D. (2018). Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *International Journal of Exercise Science*, 11(7), 503-515.
- Consumer Technology Association (2016). *Physical Activity Monitoring for Fitness Wearables: Step Counting*. In ANSI/CTA Standard (Vol. CTA-2065, pp. 1-12). Arlington VA: Consumer Technology Association: Technology & Standards Department.
- Consumer Technology Association (2018). *Physical Activity Monitoring for Heart Rate*. In CTA Standard (Vol. CTA-2065, pp. 1-21). Arlington VA: Consumer Technology Association, Technology & Standards Department.
- Coughlin, S. S., & Stewart, J. (2016). Use of Consumer Wearable Devices to Promote Physical Activity: A Review of Health Intervention Studies. *Journal of Environment and Health Science*, 2(6).
- Crouter, S., E., Schneider, P., L., & Bassett Jr., D. R. (2005). Spring-Levered versus Piezo-Electric Pedometer Accuracy in Overweight and Obese Adults. *Medicine & Science in Sports & Exercise*, 37(10), 1673-1679.
- Dondzila, C., J., Lewis, C., A., Lopez, J., R., & Parker, T., M. (2018). Congruent Accuracy of Wrist-worn Activity Trackers during Controlled and Free-living Conditions. *International Journal of Exercise Science*, 11(7), 575-584.
- El-Amrawy, F., & Nounou, M. I. (2015). Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? *Healthcare Informatics Research*, 21(4), 315-320.
- Ferris, D., P., Louie, M., & Farley, C., T. (1999). Running in the real world: adjusting leg stiffness for different surfaces. *The Royal Society*, (265), 989-994.
- Floegel, T. A., Florez-Pregonero, A., Hekler, E. B., & Buman, M. P. (2017). Validation of Consumer-Based Hip and Wrist Activity Monitors in Older Adults With Varied Ambulatory Abilities. *Journals of Gerontology Series A: Biological Sciences and Medical Sciences* 72(2), 229-236.
- Frishberg, B. A. (1983). An analysis of overground and treadmill sprinting. *Medicine and Science in Sports and Exercise*, 15(6), 478-485.
- Georgiou, K., Larentzakis, A. V., Khamis, N. N., Alshai-bani, G. I., Alaska, Y. A., & Giallafos, E. J. (2018). Can Wearable Devices Accurately Measure Heart Rate Variability? A Systematic Review. *Folia Medica (Plovdiv)*, 60(1), 7-20.
- Gillinov, S., Etiwy, M., Wang, R., Blackburn, G., Phelan, D., Gillinov, A. M., Desai, M. Y. (2017). Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Medicine and Science in Sports and Exercise*, 49(8), 1697-1703.
- How does a Polar Training Computer measure heart rate? (2018). Retrieved from <https://support.polar.com/us-en/>



- support/how-does-a-polar-training-computer-measure-heart-rate?product\_id=45200&category=faqs.
- Huang, Y., Xu, J., Yu, B., & Shull, P. B. (2016). Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. *Gait & Posture*, *48*, 36-41.
- Jones, A. M., & Doust, J. H. (1996). A 1% treadmill grade most accurately reflects the energetic cost of outdoor running. *Journal of Sports Sciences*, *14*(4), 321-327.
- Kisilevsky, B. S., & Brown, C. A. (2016). Comparison of fetal and maternal heart rate measures using electrocardiographic and cardiocographic methods. *Infant Behavior and Development*, *42*, 142-151.
- Kooiman, T. J., Dontje, M. L., Sprenger, S. R., Krijnen, W. P., van der Schans, C. P., & de Groot, M. (2015). Reliability and validity of ten consumer activity trackers. *BMC Sports Science, Medicine and Rehabilitation*, *7*, 24.
- Lamont, R. M., Daniel, H. L., Payne, C. L., & Brauer, S. G. (2018). Accuracy of wearable physical activity trackers in people with Parkinson's disease. *Gait & Posture*, *63*, 104-108.
- Liguori, G., Dweyer, G., B., & Fitts, T., C. (2014). *Resources for the Health Fitness Specialist* (1<sup>st</sup> ed.). Philadelphia PA: Lippincott, Williams, and Wilkins.
- Maeda, Y., Seaman, K., & Tamura, T. (2010). Relationship between measurement site and motion artifact in wearable reflected photoplethysmography. *Journal of Medical Systems*, *35*(5), 969-976.
- Maeda, Y., Sekine, M., & Tamura, T. (2011). The advantages of wearable green reflected photoplethysmography. *Journal of Medical Systems*, *35*(5), 829-834.
- Melanson, E. L., Knoll, J. R., Bell, M. L., Donahoo, W. T., Hill, J. O., Nysse, L. J., Levine, J. A. (2004). Commercially available pedometers: considerations for accurate step counting. *Preventive Medicine*, *39*(2), 361-368.
- Modave, F., Guo, Y., Bian, J., Gurka, M. J., Parish, A., Smith, M. D., Buford, T. W. (2017). Mobile Device Accuracy for Step Counting Across Age Groups. *JMIR Mhealth Uhealth*, *5*(6), e88.
- Montes, J., & Navalta, J. W. (2019). Reliability of the Polar T31 Uncoded Heart Rate Monitor in Free Motion and Treadmill Activities. *International Journal of Exercise Science*, *12*(4), 69-76.
- Montes, J., Young, J. C., Tandy, R., & Navalta, J. W. (2017). Fitbit Flex: Energy Expenditure and Step Count Evaluation. *Journal of Exercise Physiology online*, *20*(5), 152-159.
- Montes, J., Young, J. C., Tandy, R., & Navalta, J. W. (2018). Reliability and Validation of the Hexoskin Wearable Bio-Collection Device During Walking Conditions. *International Journal of Exercise Science*, *11*(7), 808-816.
- Murray, M. P., Spurr, G. B., Sepic, S. B., Gardner, G. M., & Mollinger, L. A. (1985). Treadmill vs. floor walking: kinematics, electromyogram, and heart rate. *Journal of Applied Physiology*, *59*(1), 87-91.
- Navalta, J. W., Montes, J., Bodell, N. G., Aguilar, C. D., Lujan, A., Guzman, G., & Kam, B. K. (2018). Wearable Device Validity in Determining Step Count During Hiking and Trail Running. *Journal for the Measurement of Physical Behaviour*, *1*, 86-93.
- Pickering, T. G., Gerin, W., & Schwartz, A. R. (2002). What is the white-coat effect and how should it be measured? *Blood Pressure Monitoring*, *7*(6), 293-300.
- Riley, P. O., Dicharry, J., Franz, J., Della Croce, U., Wilder, R. P., & Kerrigan, D. C. (2008). A kinematics and kinetic comparison of overground and treadmill running. *Medicine and Science in Sports and Exercise*, *40*(6), 1093-1100.
- Schache, A. G., Blanch, P. D., Rath, D. A., Wrigley, T. V., Starr, R., & Bennell, K. L. (2001). A comparison of overground and treadmill running for measuring the three-dimensional kinematics of the lumbo-pelvic-hip complex. *Clinical Biomechanics*, *16*(8), 667-680.
- Schneider, P. L., Crouter, S., & Bassett, D. R. (2004). Pedometer measures of free-living physical activity: comparison of 13 models. *Medicine and Science in Sports and Exercise*, *36*(2), 331-335.
- Shcherbina, A., Mattsson, C. M., Waggott, D., Salisbury, H., Christle, J. W., Hastie, T., Ashley, E. A. (2017). Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *Journal of Personalized Medicine*, *7*(2).
- Shepherd, E. F., Toloza, E., McClung, C. D., & Schmalzried, T. P. (1999). Step activity monitor: increased accuracy in quantifying ambulatory activity. *Journal of Orthopaedic Research*, *17*(5), 703-708.
- Simpson, L. A., Eng, J. J., Klassen, T. D., Lim, S. B., Louie, D. R., Parappilly, B., Zbogor, D. (2015). Capturing Step Counts at Slow Walking Speeds in Older Adults: Comparison of Ankle and Waist Placement of Measuring Device. *Journal of Rehabilitation Medicine*, *47*(9), 830-835.
- Stahl, S. E., An, H. S., Dinkel, D. M., Noble, J. M., & Lee, J. M. (2016). How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport & Exercise Medicine*, *2*(1), e000106.
- Statista. (2018a, 06/01/2018). Forecast wearables unit shipments worldwide from 2014 to 2022 (in millions). Retrieved from <https://www.statista.com/statistics/437871/wearables-worldwide-shipments/>
- Statista. (2018b, 09/01/2017). Wearable device sales revenue worldwide from 2016 to 2022 (in billion U.S. dollars). Retrieved from <https://www.statista.com/statistics/610447/wearable-device-revenue-worldwide/>
- Thiebaud, R. S., Funk, M. D., Patton, J. C., Massey, B. L., Shay, T. E., Schmidt, M. G., & Giovannitti, N. (2018). Validity of wrist-worn consumer products to measure heart rate and energy expenditure. *Digital Health*, *4*.
- Tudor-Locke, C., Barreira, T. V., & Schuna, J. M., Jr. (2015). Comparison of step outputs for waist and wrist accelerometer attachment sites. *Medicine and Science in Sports and Exercise*, *47*(4), 839-842.
- Wahl, Y., Duking, P., Droszez, A., Wahl, P., & Mester, J. (2017). Criterion-Validity of Commercially Available Physical Activity Tracker to Estimate Step Count, Covered Distance and Energy Expenditure during Sports Conditions. *Frontiers in Physiology*, *8*, 725.
- Wallen, M. P., Gomersall, S. R., Keating, S. E., Wisloff, U., & Coombes, J. S. (2016). Accuracy of Heart Rate

- Watches: Implications for Weight Management. *PLoS One*, 11(5), e0154420.
- Wilson, T. E., & Crandall, C. G. (2011). Effect of thermal stress on cardiac function. *Exercise and Sport Sciences Reviews*, 39(1), 12-17.
- Xie, J., Wen, D., Liang, L., Jia, Y., Gao, L., & Lei, J. (2018). Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. *Journal of Medical Internet Research*, 20(4), 1-1.